TI 2020-003/VIII
Tinbergen Institute Discussion Paper

Leaving the Tub: the Nature and Dynamics of Hypercongestion in a Bathtub Model with a Restricted Downstream Exit

Yue Bao[1]
Erik T. Verhoef[1]
Paul Koster[1]

[1] Department of Spatial Economics, VU University Amsterdam, Amsterdam, the Netherlands

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at https://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Leaving the Tub: the Nature and Dynamics of Hypercongestion in a Bathtub Model with a Restricted Downstream Exit

Yue Bao,[a] Erik T. Verhoef,[a,b] Paul Koster[a,b]

[a]Department of Spatial Economics, VU University Amsterdam, Amsterdam, the Netherlands

[b]Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, the Netherlands

Abstract: Hypercongestion is the situation where a certain traffic flow occurs at a combination of low speed and high density, and a more favorable combination of these could produce the same flow. The macroscopic fundamental diagram (MFD) allows for such hypercongestion, but does not explicitly describe the dynamic process leading up to hypercongestion. Earlier studies of hypercongestion on single links have, however, confirms that such dynamic processes are important to consider. The bathtub model is one class of model that can be used to investigate how hypercongestion can arise in urban areas, when drivers can choose their departure times. This paper investigates equilibrium outcomes and user costs under the realistic assumption that there is finite capacity to exit the bathtub, without which it would be hard to explain why hypercongestion would not dissolve through shockwaves originating from the bathtub exit. We find that when the exit capacity of the bathtub is lower than the equilibrium exit flow from the bathtub, no additional inefficiencies arise due to hypercongestion. However, when the exit capacity is higher than the equilibrium exit flows from the bathtub inefficiencies do occur. The implication of this result is that a restricted downstream capacity regulating hypercongestion may not be necessary because time gains in the bathtub are lost in the queue at the exit.

Keywords: Road traffic congestion; flow congestion; bottleneck model; bathtub model; hypercongestion; macroscopic fundamental diagram

1. Introduction

Traffic congestion arises because of limited road capacities and concentrated demand in space and time of the day. This leads to travel time losses, scheduling costs, and may enhance environmental externalities such as air pollution and traffic accidents. When travelers make a trip to an urban region, they usually face two types of congestion: bottleneck queueing behind bridges or tunnels and dynamic flow congestion in urban areas. Bottleneck congestion was first studied from a dynamic equilibrium perspective by Vickrey (1969), later elaborated upon in papers by Arnott et al. (1990; 1993). Dynamic equilibrium in these models arises because the capacity of the bottleneck is limited and travelers make trade-offs between travel time losses and scheduling disutilities from arriving earlier or later at their destination than preferred. In the bottleneck model, the only travel time loss, above the free flow travel time, is the waiting time at the queue before the bottleneck. This feature of the bottleneck model leads to convenient closed-form results for the traffic equilibrium and optimum, which makes this stylized bottleneck model attractive for the economic analysis of dynamic equilibrium (Arnott et al. 1993).

However, additional time losses can result from flow congestion on roads other than bottlenecks, where equilibrium speeds could fall below free-flow speed levels before maximum flow capacity is reached. In flow congestion models, the dynamic travel time function provides a structural relationship between capacity, the number or density of travelers at a certain time instant, and the speed or travel time. Many such functions display delays for use below capacity. For example, Henderson (1974) applies a formulation where the travel time for a single road is determined by the flow at the road's entrance when departing from home, whereas Chu (1995) uses a specification where travel time is determined by the flow at the road's exit when arriving at the destination. Mun (1994) and Mun (1999) divided a one-link road into two parts of endogenous lengths: a part without a queue, followed by a traffic jam part caused by a downstream bottleneck. Others have investigated the impacts of queue spillback and capacity drops at the bottleneck on the upstream flow (e.g., Daganzo, 1998; Nie & Zhang, 2008; Ma et al., 2017; Yuan et al, 2017; Baer et al., 2019).

In recent years, there have been new developments in the modelling of peak congestion in cities (e.g., Geroliminis & Levinson, 2009; Arnott, 2013; Fosgerau & Small, 2013; Fosgerau, 2015; Daganzo & Lehe, 2015; Lamotte & Geroliminis, 2018), arising from empirical observations about congestion in urban areas. An important empirical insight is the observation of a stable relationship between average equilibrium speed and average equilibrium density at the level of an entire urban area (Geroliminis & Daganzo, 2008). This relationship is called the Macroscopic Fundamental Diagram (MFD) in order to distinguish it from the familiar link-based fundamental diagram of traffic congestion that depicts the inverted U-shape relation between density and flow. The MFD arises as the combined result from speed falling monotonously with density, and flow equaling the product of density and speed. A similar stable relationship between traffic speed, density and flow for an entire urban area also underlies models of "bathtub congestion" proposed in various recent transport economic models (Small & Chu, 2003; Arnott, 2013; Fosgerau, 2015). The usual explanation for the parallel with a bathtub is that for an urban area, the process of drivers entering (leaving) the traffic network is like the inflow (outflow) of water into (out of) bathtub, where traffic density would correspond to the level of water in the bathtub (Fosgerau, 2015). In the bathtub model, instantaneous traffic conditions are assumed to be homogeneous over continuous space, making it possible to avoid the explicit modelling of route choices. Furthermore, this setup avoids complications arising from continuous-time – continuous place congestion modelling, as in for example car-following models (Verhoef, 2001; 2003) and hydro-dynamic models (Newell, 1988). In contrast to the bottleneck model, the bathtub model allows for hypercongestion, implying that equilibrium travel times can decrease in equilibrium traffic flows.

The MFD describes such hypercongestion, by depicting that a certain flow can be consistent with two equilibrium speeds, the lower of which being the hypercongested one. In this paper we will explore what happens in this sort of framework when the persistence of hypercongestion in the bathtub is explained from a limitation on the exit capacity. There is good reason to consider this explicitly: earlier work on hypercongestion confirmed that without a limited exit capacity, hypercongestion would dissolve over time, the dissolution starting from the exit and subsequently moving upstream (Verhoef, 2003). Whereas the earlier bathtub models impose, implicitly, that traffic conditions be uniform over space so that such dissolution cannot occur by construction so that no downstream capacity limitation is needed to create hypercongestion, we explicitly model this downstream capacity limitation, by assuming that there is an exit bottleneck with finite capacity. We will see that, when the exit capacity of the bathtub is limited, this does not imply that moving from the hypercongested situation (B2 in Fig 1) to the situation with flow congestion (B1 in Fig 1) will necessarily lead to a better performance of the system in terms of user costs and therewith overall welfare. This is in particular relevant if ,despite the lower hypercongested flow in the bathtub, the exit flows from the bathtub remain constant because that exit has a fixed capacity and functions as a bottleneck. A higher speed in the bathtub may then simply mean an earlier joining of the queue at the exit, so that travel time gains in the bathtub are nullified by increased queuing time at the exit, and the moment of trip completion remains unaltered, being determined by the exit's capacity. When departure time choices of travelers are endogenous, the equilibrium duration of the peak period will depend on the downstream capacity when it operates at full capacity throughout the peak. Only when the outflow from the bathtub drops below the exit capacity, the equilibrium duration of the peak may be affected by the flow in the bathtub, and benefits can be obtained by moving from the hypercongested situation (A2 in Fig 1) to the normally congested range (A1 in Fig 1). This implies that if the exit continues to operate at full capacity, it does not matter whether speed in the bathtub varies between $\breve{v}$ and $\hat{v}$ (with hypercongestion), or between v^* and $\hat{v}$ (without hypercongestion). The duration of the peak period is then always equal to the ratio between demand and the downstream capacity: N/c_a, and travel delays in the bathtub are fully substituted by travel delays at the exit queue. Only when speed further increases from $\hat{v}$ (normal congestion) or further decreases from $\breve{v}$ (hypercongestion), exit capacity waste may occur if that deviation is sufficiently long-live to make an early exit queue fully dissipate, and this would then result in a longer peak period. This paper further investigates this intuition analytically, and analyses these two types of potential equilibrium outcomes in more detail.

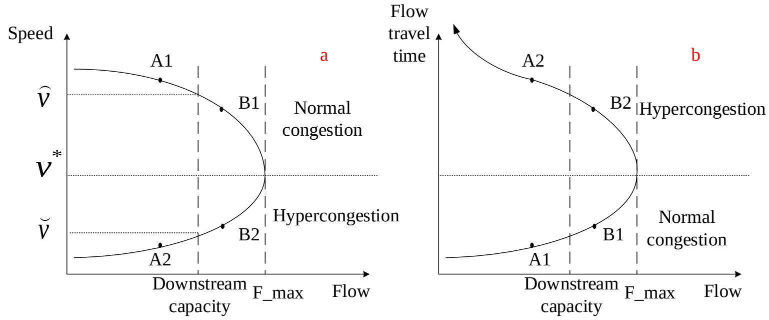


Fig 1. Hypercongestion and downstream capacity

For downtown traffic, an interpretation of our restricted exit capacity can be parking facilities. Drivers leave the homogeneous traffic stream and enter the parking areas, with limited (entrance) capacity. Each parking area then functions as a pure bottleneck in our models, meaning that there is no time loss associated with the use of the facility when it has no queue, and the arrival rate at the facility does not exceed its capacity, while in all other cases a queue grows (or shrinks, with negative growth) over time, at a rate equal to the difference between the arrival rate of new users at the back of the queue and the capacity of the exit. It should be noted that we only consider the limited entrance capacity, i.e., limited flow capacity, of the parking area, but do not involve the stock capacity inside the parking areas, to simplify the argument of the paper. We also wish to maintain the analytical advantages of homogenous traffic conditions in the bathtub, and for that reason ignore spillbacks from the exit queues into the bathtub. So, all queuing congestion at the exit takes place in a way such that it does not affect traffic conditions in the bathtub. For a more detailed analysis of drivers' cruising for parking space with a bathtub model we refer to Geroliminis (2015) and Liu & Geroliminis (2016), and for spillback we refer to Mariotte & Leclercq (2019).

We will adopt two alternative approaches to model the equilibrium traffic patterns in the bathtub, inspired by alternative formulations proposed by Arnott (2013) and Fosgerau (2015). For the first, we assume that the dynamic outflow from the bathtub at any instant is proportional to the dynamic traffic flow in the bathtub. We will refer to this as the CPI (constant proportion of inflow into the bottleneck) model in this paper. Gonzales & Daganzo (2012) called this relationship the "Network Exit Function (NEF)", because the flow exiting the network is a function of the total number of vehicles circulating in the network.

It is demonstrated by Geroliminis & Daganzo (2007) that this function can be estimated using the MFD of the network. This feature of MFD is also used by Amirgholy & Gao (2017) and Lamotte & Geroliminis (2018). The second method employs the relationship between trip length, travel time and traffic speed, and will be referred to as the FTL (fixed trip length) model. Given that trip length is equal to the integral of (time-varying) traffic speed from the starting moment to the moment of completion of the journey, the derivative of the arrival time with respect to departure time is equal to the ratio of the traffic speeds at the moment of departure and the moment of arrival (Fosgerau, 2015). This approach is also termed as "trip-based" in literature (Mariotte et al., 2017).

Hypercongestion is verified to exist for both model formulations, where the duration of hypercongestion naturally depends both on the properties of the network and the assumed demand parameters. However, we will be more elaborate than this as in the context of our model, two types of hypercongestion can be distinguished. The first is "bathtub-speed" hypercongestion, which arises when traffic density in the bathtub exceeds the value consistent with the maximum possible flow in the bathtub (see Fig. 1, B1). The inflow into the queue then naturally falls below the maximum possible inflow, determined by the maximum flow that can occur in the bathtub. The second form is "system" hypercongestion, and this arises when a sufficiently extended period of bathtub-speed hypercongestion makes queuing at the bottleneck disappear altogether, and subsequently causes outflow from the bottleneck to fall below the bottleneck capacity (see Fig. 1, A1) for a certain period where the bathtub is heavily congested. While bathtub-speed hypercongestion only involves zero-sum tradeoffs between travel time in the bathtub and queuing time at the bottleneck, system hypercongestion results in an inefficiently long equilibrium peak period, and corresponding higher equilibrium travel cost. It is demonstrated in the numerical examples that bathtub-speed hypercongestion is a necessary but not a sufficient condition for system hypercongestion to occur.

One interesting and surprising finding of this paper is that the addition of a bottleneck at the exit of the bathtub makes the proposed model easier to be solved analytically and results in closed-form expressions for generalized user costs and equilibrium peak duration. Earlier authors have solved the model using numerical simulation of delay-differential equations or using additional assumptions. For example, Geroliminis & Levinson (2009) adopted the first-in-first-out (FIFO) principle where travel time is determined by the cumulative number of cars in the system when one arrives. Arnott (2013) assumed that drivers' remaining trip lengths at any time in the bathtub follow an exponential distribution, and the expected trip length is fixed and homogenous across drivers. Fosgerau (2015) shows that the model can be solved numerically when assuming heterogeneous trip lengths in an equilibrium with "regular sorting", meaning that drivers with longer trip lengths will depart earlier and arrive later in equilibrium. Amirgholy & Gao (2017) adopted exogenous departure rate functions and shows differences across the various "approximating models" are very small. Arnott & Buli (2018) developed a computational method to obtain numerical solutions, assuming smooth preferences.

In contrast, the proposed model of this paper has mathematical features that allow for analytical solutions. At the same time, we add a downstream bottleneck which as argued, is needed to maintain hypercongestion once it is allowed for traffic conditions to vary over space near an exit. We show that the resulting equilibrium departure profile has a similar shape as in the conventional bottleneck model and is independent of the traffic situation in the bathtub area, reflecting the perfect substitution between delays in the bathtub and in the exit queue. Using a numerical example, we show first that the two models can lead to different predictions of the dynamic equilibrium travel time in the bathtub and the exit queue for the two approaches (the CPI model and the FTL model). However, the differences between the equilibrium generalized user cost and the equilibrium peak period for the two approaches mentioned above are negligible. The conceptual differences between the two approaches are well-defined and there can be reasons to prefer, depending on the situation studied, the one formulation over the other, but numerically the differences in user costs are very small. This is in fact consistent with the findings in Amirgholy & Gao (2017), who find that aggregate travel costs are not very different across the various "approximating models".

The remainder of the paper is organized as follows. The next section formulates the proposed models and compares the differences between the two approaches in the literature. Section 3 presents a numerical example to illustrate the models and analyzes the impacts of the network properties and the travel demand properties. The final section concludes the paper and discusses possible extensions for future research.

Section 2. The bathtub model with a bottleneck at the exit

2.1 The bathtub model without a bottleneck at the exit

The bathtub model assumes that the urban area is homogeneous and results in equilibrium traffic speeds, densities, and flows that are identical over space because drivers can continuously adjust their route to avoid more congested parts of the road network (Fosgerau, 2015). Flow congestion in the bathtub area follows a stable relationship between traffic density and traffic speed. For this paper, this relationship is assumed to be linear and given by (Greenshields et al., 1935):

$$v\left(k(t)\right)=v_f\left(1-\lambda k(t)\right), \tag{1}$$

where v_f is the free-flow traffic speed and $\lambda = 1/k_j$ is defined as the reciprocal of the jam density, k_j, which is the value of traffic density that results in zero traffic speeds. The traffic density $k(t)$ is the total number of cars in the bathtub at time t, and is equal to the difference between cumulative departures from home, $D(t)$, and the cumulative exits from the bathtub (or the aggregate inflow to the bottleneck) at time t, $A(t)$:

$$k(t)=D(t)-A(t). \tag{2}$$

Denote $d(t)=D'(t)$ and $a(t)=A'(t)$ as the departure rate from home (inflow into the bathtub) and the arrival rate at the bottleneck (outflow from the bathtub into the bottleneck queue), respectively. The evolution of traffic density in the bathtub is then given by the difference in the departure rate from home and the arrival rate at the bottleneck:

$$k'(t)=d(t)-a(t). \tag{3}$$

Besides the traffic speed, the travel time in the bathtub is also determined by the trip length in the bathtub, L. The relationship between the trip length and the departure time is logically equal to the integral over the speed from the time of departure until the time of arrival at the bottleneck queue, $s(t)$:

$$L = \int_{t}^{s(t)} v\big(k(\omega)\big) d\omega, \tag{4}$$

where $s(t)$ denotes the arrival time at the bottleneck queue for a driver departing at time t. Then, assuming first-in-first-out principles, the cumulative departures from home and the cumulative arrivals at the bottleneck are related as:

$$D(t) = A\big(s(t)\big), \tag{5}$$

which means that the cumulative number of drivers having departed from home at time t is equal to the cumulative number of drivers having arrived at the bottleneck at time $s(t)$.

Unfortunately, this bathtub model cannot be solved without making further assumptions. The difficulty is that, in order to obtain the equilibrium departure and arrival schedules, we need to know drivers' arrival time at the destination, while this arrival time depends on the traffic speed and density at every instant in the bathtub before arriving at the destination. However, the instantaneous traffic speed and density are determined by the departure and arrival schedules. There are two approaches available in the literature to make further progress. First, Arnott (2013) adopted a proportional relationship between the outflow from the bathtub and the flow in the bathtub, which is obtained through an assumption on probabilistic arrivals. The relationship is consistent with the NEF in Gonzales & Daganzo (2012), which describes "the flow of vehicles exiting the network as a function of the total number of vehicles circulating in the network". The same assumption is adopted later in Amirgholy & Gao (2017).

The second approach to obtain an equilibrium solution was introduced by Fosgerau (2015). He considers heterogeneous trip lengths, l, that follow a distribution in the population. Here, individual drivers know their actual trip length before leaving home. Fosgerau (2015) shows that in some but not all cases travelers sort themselves in equilibrium, where drivers with a longer trip length will depart earlier and arrive later in equilibrium. Denote departure time and arrival time as functions of trip length. From condition (4), the derivative of arrival time with respect to trip length is obtained

$$s'(l)=\frac{1}{v(s(l))}+\frac{v(t(l))}{v(s(l))}t'(l). \tag{6}$$

Inserting the above condition (6) into the first-order condition of a dynamic equilibrium, namely that the generalized trip cost should be constant over time for homogenous drivers as long as they depart and arrive, drivers' travel time only depends on the speed-density relationship (1), and on the distribution of trip lengths. Interestingly, the addition of heterogeneous trip lengths makes it possible to solve the model.

2.2 Comparing one-directional road systems and spatial homogeneous road systems

In a one-directional road system, downstream capacity may have an important impact on the congestion pattern in upstream traffic (e.g., Verhoef, 2003; Arnott and Inci, 2006). Mun (1994; 1996) modelled such a one-link road network with a downstream bottleneck. Traffic forms a queue behind the bottleneck, which affects the upstream traffic in his model. Travel time consists of two parts: travel time on the road without queue, and the travel delay caused by queue. Both are monotonously increasing with the traffic flow. Therefore travel time always increases with traffic flow, and the backward-bending part of travel costs as depicted in Fig. 1 does not appear, so hypercongestion does not occur in this model.

In contrast, Verhoef (2003) used a car-following model to show how traffic speed evolves when approaching a downstream bottleneck. Given a sufficiently large travel demand, he shows that hypercongestion occurs in a dynamic equilibrium in a queue resulting from a downstream bottleneck. That bottleneck leads to an increase in density, and a decrease in speed, first close to that bottleneck and subsequently also at places further upstream. This involves speeds and densities from the hypercongested range of the speed-flow relation, so hypercongested speed occur. At the same time, for a downstream capacity that is larger than upstream capacity, any initially assumed hypercongestion will dissolve over time. A higher exit capacity results in a higher traffic speed, which makes the density decrease, which makes speeds increase further, etc.; to a level that is no longer hypercongested.

Different from a one-directional road network, a larger exiting capacity cannot dissolve the hypercongestion when a spatial homogenous bathtub is assumed. Without constraints on the inflow into the bathtub, an exiting capacity that is larger than the maximum flow in the bathtub would always keep the outflow from the bathtub at the same level as the flow in the bathtub, simply because spatial homogeneity prevails by construction. At the same time, there is no limit on the inflow into the bathtub, so that density can indeed rise to hypercongested levels, irrespective of the outflow capacity. The unlimited inflow potential, and the independence of traffic conditions from outflow capacity, make the dynamic mechanism behind the build-up of hypercongestion in the bathtub entirely different from those in one-directional systems. As the assumption that hypercongestion cannot dissolve from the exit is less unrealistic if exit capacity is finite, we include this feature in our model, to see how it would affect bathtub traffic conditions over the peak.

2.3 The bathtub model with a bottleneck at the exit

The key difference of our model with earlier bathtub models in the literature is that we consider a stylized bathtub area with a bottleneck at the exit. We employ two alternative model formulations with different solutions. We start with the assumption of instantaneous outflow of the bathtub related to flow, the CPI model. The model is inspired by the model of Arnott (2013), but does not require arrivals to be probabilistic. Our second model, the FTL model, is motivated by Fosgerau (2015), and makes use of the relationship between trip length and traffic speed as given by Eq. (4). The difference with the CPI Model is that the trip length in the bathtub is assumed to be fixed for all drivers (Fosgerau, 2015).

We consider commuting trips in which drivers need to drive through a city (the bathtub area) to their work. Consistent with the assumption of spatial homogeneity of traffic conditions in a bathtub, we assume that drivers' destinations are uniformly distributed in the bathtub. Thus, exits from the bathtub are distributed uniformly over space, in such a fine resolution that they induce no spatial inhomogeneities in traffic conditions. The description most closely representing our model would be a spatial continuum of exits, which we will approximate by letting the number of exits, E, go to infinity. To maintain spatial homogeneity, we assume that these exits all have the same capacity, and are equally spaced, so that exit capacity per unit of space in the bathtub is also constant over space. Denoting the aggregate outflow from the bathtub into all exit queues jointly as $\overline{A}(t)$, the arrival rate of new users at each bottleneck – or at the tail of its queue – is, under spatial homogeneity, $\overline{A}(t)/E$. Denoting the aggregate exit capacity from the bathtub as $\overline{c}_a$, the capacity per bottleneck will be $\overline{c}_a/E$.

Now observe that for each of these bottlenecks, the queuing time at any moment is a function of an integral, over time, of earlier queue growth rates $\left(\overline{A}(t)-\overline{c}_a\right)/E$, divided by the bottleneck capacity $\overline{c}_a/E$. A consequence is that we can multiply any single bottleneck's queue growth rates and capacity by E and find the same time pattern of queuing delays for that bottleneck. In other words, we can treat the continuum of exits as if they form a single bottleneck with capacity c_a, with an aggregate inflow of $A(t)$. That is what we will do in what follows.

Concerns over spatial inhomogeneities in traffic conditions, that a localized single exit point of outflow would undoubtedly induce, vanish when E goes to infinity while keeping the exits and their capacities distributed uniformly over space. But we can maintain the relatively simple dynamics that a single bottleneck brings about. The assumed uniform distribution of exit capacities over the bathtub is of course a simplification, but it seems the only description that is consistent with the spatial homogeneity of traffic conditions that characterizes bathtub congestion models in the first place. In particular, spatially homogenous time-varying traffic conditions in the bathtub can only be supported as a dynamic equilibrium if also time-varying exit conditions are homogeneous over space.

Finally, note that we do not explicitly model inflow into the bathtub. Also here, the assumption of spatially homogeneous traffic conditions dictate that the implicit assumption is that the inflow into the bathtub is spatially homogeneous. One way to visualize this, is to imagine cars "raining down" or "popping up" into the bathtub in a spatially homogeneous fashion. Certainly, this is not a very realistic picture; it is a simplifying assumption that one has to make in order to be able to invoke the analytically highly convenient assumption of spatially homogeneous traffic conditions, allowing the use of time-varying but non-spatial variables (hence single-valued at any moment in time) for the fundamental traffic conditions flow, speed and density.

In what follows we will thus treat the exit of the bathtub as a single bottleneck with an aggregate attempted time-varying inflow of $A(t)$ and a capacity of c_a, but do this keeping in mind that it produces queuing times that would apply for a continuum of spatially homogenous bottlenecks with the same aggregate inflow and capacity. We therefore formulate a mathematical model that considers a spatially homogeneous bathtub area with a single bottleneck at its exit. Commuters depart from home, encounter the traffic flow congestion in the city/bathtub area, and try to minimize congestion delays by continuously adapting route choices and searching a proper parking lot, which eventually results in homogenous equilibrium traffic speeds, flows and densities over space in the bathtub, including the exits. Then they arrive at the exit and face a drop of capacity at the bottleneck (or the entrance to the parking lot), which results in a build-up of the queue. That queue is assumed to be vertical in the sense that it does not occupy space in the bathtub, and does not enter the expression for traffic density in the bathtub. After passing the bottleneck, they arrive at their office with scheduling costs due to not arriving at their preferred arrival time.

We first assume that the capacities of the bathtub and the bottleneck are at such levels that the bottleneck is always fully utilized from the first driver arriving at the bottleneck until the last driver arriving at the bottleneck. This assumption will be relaxed in Section 4 to show different types of hypercongestion. The queuing time for a traveler who departs from home at time t and arrives at the bottleneck at time $s(t)$ is given by:

$$T_q\left(s(t)\right)=\frac{A\left(s(t)\right)-c_a\left(s(t)-t_q\right)}{c_a}, \tag{7}$$

where c_a is the capacity of the bottleneck and t_q is the arrival time at the bottleneck (and also at destination) of the very first driver departed, i.e., $t_q=s(t_s)$. For a driver departing at time t, the arrival time at the destination is given by:

$$\psi(t)=s(t)+T_q\left(s(t)\right)=\frac{A\left(s(t)\right)}{c_a}+t_q=\frac{D(t)}{c_a}+t_q\,, \tag{8}$$

which satisfies:

$$\psi'(t)=s'(t)+T_q'\left(s(t)\right)=\frac{d(t)}{c_a}. \tag{9}$$

Denote t^* as the drivers' preferred arrival time at the destination. Drivers' user cost in monetary units is assumed to be given by conventional expressions for "$\alpha-\beta-\gamma$" preferences:

$$p(t)=\alpha\left(s(t)+T_q\left(s(t)\right)-t\right)+\max\left\{\beta\left(t^*-s(t)-T_q\left(s(t)\right)\right),\gamma\left(s(t)+T_q\left(s(t)\right)-t^*\right)\right\}, \tag{10}$$

where α denote drivers' value of time, β and γ are the unit schedule-early and the unit schedule-late cost, respectively. The first-order condition for 'drivers' optimal departure time choice is given by:

$$p'(t)=\begin{cases}\alpha\left(\dfrac{d(t)}{c_a}-1\right)-\beta\dfrac{d(t)}{c_a}, t_s\le t\le\tilde{t}\\ \alpha\left(\dfrac{d(t)}{c_a}-1\right)+\gamma\dfrac{d(t)}{c_a}, \tilde{t}\le t\le t_e\end{cases}=0\,, \tag{11}$$

where t_s and t_e denote the times of the first departure and the last departure from home, and $\tilde{t}$ is the departure time from home of the drivers who arrive at the destination at the preferred arrival time t^*. Solving Eq. (11) for the departure rate yields:

$$d(t)=\begin{cases}\phi_1, t_s\le t\le\tilde{t}\\ \phi_2, \tilde{t}<t\le t_e\end{cases}, \tag{12}$$

where $\phi_1=\dfrac{\alpha}{\alpha-\beta}c_a$, $\phi_2=\dfrac{\alpha}{\alpha+\gamma}c_a$. It implies the equilibrium departure rate is the same as in the conventional bottleneck model and is first higher than the bottleneck capacity and then lower than the bottleneck capacity (Vickrey, 1969). This is because when the bottleneck at the exit of the bathtub is always fully utilized and based on the first-in-first-out principles, drivers must queue at the bottleneck in the order of their departure time. Thus, the arrival time at the destination is determined by how many drivers departing before them, the capacity of the bottleneck and the arrival time of the first departure but not their travel time in the bathtub. It means that basically we have a standard bottleneck model, where for an individual's total travel delay cost, it is only the sum of bathtub travel time and bottleneck queuing time that matters. The cumulative departures are given by:

$$D(t) = \begin{cases} \phi_1(t - t_s), t_s \le t \le \tilde{t} \\ \phi_2(t - t_s) + \phi_3, \tilde{t} < t \le t_e \end{cases}, \tag{13}$$

where $\phi_3 = \dfrac{\gamma + \beta}{\alpha + \gamma} c_a \left(t^* - t_q\right)$.

As there are no arrivals at the bottleneck before the first driver arrives there, the average density in the bathtub is equal to the cumulative departures for $t_s \le t < t_q$, implying (for a more detailed analysis on the traffic density on the shoulder of the peak period we refer to Appendix B):

$$k(t) = D(t) = \phi_1(t - t_s), t_s < t \le t_q. \tag{14}$$

Consistent with this, there are no departures after t_e, while arrivals at the bottleneck still occur. Therefore, for $t_e < t \le s(t_e)$, the density in the bathtub is equal to the total number of drivers minus the cumulative arrivals at the bottleneck:

$$k(t) = N - A(t), t_e < t \le s(t_e). \tag{15}$$

From the speed-density relationship in Eq. (1) and the traffic density function in Eq. (14), the first driver's travel time is given by

$$t_q = t_s + \frac{1}{\lambda \phi_1}\left(1 - \sqrt{1 - 2\lambda \phi_1 \frac{L}{v_f}}\right) \tag{16}$$

In other words, the first driver already chooses the same speed as his/her followers when he/she travels. Again, the assumption has the benefit of consistency of assumptions, but the disadvantage of being unrealistic: the very first driver is slowed down because of upstream density; i.e., because of traffic conditions behind. For the last driver, however, the travel time can only be determined when fully solving the models. In what follows, we will consider two ways to model the traffic situation in the bathtub, given this knowledge of the equilibrium departure pattern.

2.4 Model I: Constant Proportion of Inflow into the bottleneck (CPI)

To allow us to consider a conventional bathtub area but with a bottleneck at the exit, and maintain the standard assumption that traffic density, speed, and flow are homogeneous over space in the bathtub, we should assume that the queue is a standard spaceless, vertical "Vickrey" queue (Vickrey, 1969). For the CPI model, the arrival rate at the bottleneck queue (the instantaneous outflow of the bathtub) should be proportional to the flow in the bathtub, as for the empirical NEF function, resulting in:

$$a(t) = \eta k(t) v(k(t)), \tag{17}$$

where η is the ratio between the outflow from the bathtub (or inflow into the bottleneck queue) and the flow in the bathtub, which is related to the size of the bathtub and the availability and size of the ubiquitous exit facilities (see Page 14). Substituting condition (17) into condition (3) yields:

$$k'(t) = d(t) - a(t) = \begin{cases} d(t), t_s \le t \le t_q \\ d(t) - \eta k(t) v(k(t)), t_q \le t \le t_e \end{cases}. \tag{18}$$

The arrival rate into the queue before t_q is zero, while after t_q it is equal to η times of the flow. Given the departure rate $d(t)$ in Eq. (12), the above differential equation (18) can be solved for the equilibrium traffic density, speed and flow using the boundary condition $k(t_s) = 0$. Appendix A provides more details on these derivations.

With the equilibrium departure and arrival rates, the arrival time at the bottleneck queue can be obtained using Eq. (5) (see Appendix A). As the arrival time at the bottleneck queue can also be obtained from Eq. (4), this implies that trip lengths may be unequal for different drivers in equilibrium. Therefore, the CPI model follows Arnott et al. (2013) in that drivers do not know their exact trip length in the bathtub beforehand.

2.5 Model II: Fixed Trip Length (FTL)

The FTL model builds on the relationship between trip length, traffic speed in the bathtub, the departure time from home, and the arrival time at the bottleneck given in Eq. (4), a relationship used in Fosgerau (2015). However, here we assume a non-stochastic fixed trip length, in our case identical across users. From Eq. (4) it follows that for a given L, the derivative of $s(t)$ with respect to t is given by:

$$s'(t) = \frac{v(k(t))}{v(k(s(t)))}, \tag{19}$$

implying that the change in the arrival time at the bottleneck is determined by the ratio of speed at the moment of departure and speed at the moment of arrival at the bottleneck. Combining this equation with conditions (2) and (5) we find:

$$s'(t) = \frac{v(D(t) - A(t))}{v(D(s(t)) - D(t))}. \tag{20}$$

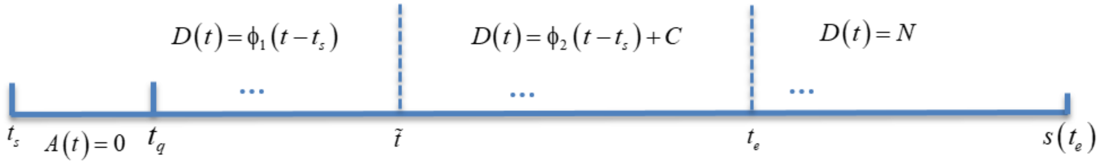
It should be noted that for t_s, the departure rate is $\phi_1 = \frac{\alpha}{\alpha - \beta} c_a$, which is larger than the capacity of the bottleneck. Therefore, the queue builds up as soon as the first group of drivers arrives at the bottleneck. Given the cumulative departures from home, $D(t)$, in Eq. (13), the above equation (20) can be solved if $A(t)$ is known. The first driver's arrival time at the bottleneck is t_q and there are no arrivals at the bottleneck before that moment. Therefore, $A(t) = 0$ for $[t_s, t_q)$, and for other moments, $A(t)$ can be obtained through Eq. (5) if we know the departure time of the driver arriving at the bottleneck at t; i.e., $s^{-1}(t)$. Therefore, we define $[t_s, t_q)$ as the first interval, $[t_q, s(t_q))$ as the second interval, $[s(t_q), s(s(t_q)))$ as the third interval unitil the last driver arrives. For the first interval, the ordinary differential equation (20) can be rewritten as:

$$s'(t) = \frac{v(D(t))}{v(D(s(t)) - D(t))}, \; t \in [t_s, t_q). \tag{21}$$

Given the traffic speed-density relationship (1), this ordinary differential equation can be solved with $s(t_s) = t_q$ for the time interval $\left[t_s, t_q\right)$. Based on Eq. (5), the cumulative arrivals $A(t)$ for time interval $\left[t_q, s(t_q)\right)$ can be obtained:

$$A(t) = D\left(s^{-1}(t)\right). \tag{22}$$

With Eqs. (13) and (22), the ordinary differential equation (20) can be numerically solved for $\left[t_q, s(t_q)\right)$ to yield the arrival time at the bottleneck queue, $s(t), t \in \left[t_q, s(t_q)\right)$ (if other types of speed-density relationship than the current linear one are adopted, such as the exponential speed-density relationship, complicated closed-form results may be obtained). Again, the cumulative arrivals, $A(t)$, for the next time interval can be obtained based on Eq. (22). Fig. 2 summarizes the method for the calculation of arrival intervals in this model. Basically, it is an iterative procedure, that links solutions of earlier arrival times to later arrival intervals. The cumulative arrivals, $A(t)$, are known for the first interval. We then obtain $s(t)$ for the first departure interval by solving the differential equation (21). With this solution to $s(t)$, we obtain $A(t)$ for the second departure interval, based on Eq. (22). Then with the solution for $A(t)$ for the second departure interval, we obtain $s(t)$ for the second departure interval by solving Eq. (20), and with the solution to $s(t)$ for the second departure interval, we obtain $A(t)$ for the third departure interval with Eq. (22). In this way, we can further calculate $s(t)$ and $A(t)$ for the remaining departure intervals until the departures end. Then, the equilibrium traffic density results from Eq. (2) and the equilibrium traffic speed results from Eq. (1).



$$s'(t)=\frac{v\big(D(t)-A(t)\big)}{v\big(D(s(t))-A(s(t))\big)}, \because A\big(s(t)\big)=D(t)$$

Fig 2. The solution procedure of the FTL Model

Although both model formulations, CPI and FTL, look reasonable, it is in general impossible for them to hold at the same time; otherwise, condition (5) is not satisfied. For the CPI model, the arrival rate at the bottleneck queue is proportional to the flow in the bathtub, which is attractive from the perspective of logical relations between flows of vehicles. But the CPI model results in time varying and stochastic trip lengths. For the FTL model, the trip length is fixed and deterministic, but the arrival rate at the bottleneck queue is not directly related to the flow in the bathtub. This has the disadvantage that the ratio between flow in the bathtub and inflow into the queue, η, in fact becomes endogenous, and time varying in this formulation. Depending on the empirical plausibility, both sets of assumptions might be more or less unattractive for the reasons mentioned above, but both can guarantee that an equilibrium can be found. Section 3 assesses numerically whether the two formulations lead to strong differences in predicted dynamic congestion patterns in the bathtub, therewith providing insight into the importance of choosing between these two model formulations.

2.6 Approximate closed-form solutions: assuming free-flow travel time for the last driver

Here we first give the peak period and the generalized travel cost when the travel time of the last driver is assumed to be free-flow travel time. This case is useful as it leads to closed-form solutions for the start and the end of the peak, as well as for the equilibrium travel costs, which can serve as approximations for true equilibrium costs. Obviously, in reality, the last driver's travel time cannot be far from the free-flow value, as otherwise it would become attractive to depart later from home and drive at higher speed. As the bottleneck is always fully utilized between $\left[t_q, s\left(t_e\right)\right]$, the total number of drivers passing the bottleneck is equal to the total demand, N, i.e.,

$$c_a\left(s\left(t_e\right)-t_q\right)=N. \tag{23}$$

Thus, the arrival time of the last driver at the destination is given by

$$s\left(t_e\right)=t_q+\frac{N}{c_a}=t_e+\frac{L}{v_f}. \tag{24}$$

Furthermore, the total number of drivers departed should also be equal to the travel demand, i.e., $D\left(t_e\right)=N$. Based on Eq. (13), the departure time of the last driver should satisfy the condition below:

$$D\left(t_e\right)=\frac{\alpha}{\alpha+\gamma}c_a\left(t_e-t_s\right)+\frac{\gamma+\beta}{\alpha+\gamma}c_a\left(t^*-t_q\right)=N \tag{25}$$

Combining the above Eqs. (24)-(25) with Eq. (16) yields

$$t_s=t^*-\frac{\gamma}{\gamma+\beta}\frac{N}{c_a}-\frac{L}{v_f}-\left(1-\frac{\alpha}{\gamma+\beta}\right)\varepsilon, \tag{26}$$

$$t_e=t^*+\frac{\beta}{\gamma+\beta}\frac{N}{c_a}-\frac{L}{v_f}-\left(1-\frac{\alpha}{\gamma+\beta}\right)\varepsilon, \tag{27}$$

where $\varepsilon=\left(1-\sqrt{1-2\lambda\phi_1 L/v_f}\right)\Big/\lambda\phi_1-L/v_f$ is the extra travel time above the free-flow travel time which will always be positive. The resulting equilibrium generalized travel cost for this case are then given by:

$$u\left(t\right)=\alpha\frac{L}{v_f}+\frac{\gamma\beta}{\gamma+\beta}\frac{N}{c_a}+\frac{\alpha\gamma}{\gamma+\beta}\varepsilon, \tag{28}$$

where the last part of this equation is the extra travel time the first driver needs above free-flow travel time multiplied by $\frac{\alpha\gamma}{\gamma+\beta}$. Compared to the conventional bottleneck model, it therefore has extra equilibrium travel cost caused by the increase of the first driver's travel time which is equal to the extra travel time cost multiplied by $\frac{\gamma}{\gamma+\beta}$. In other words, it implies that when the first driver's travel time increases by ε, the equilibrium generalized travel cost increases by $\frac{\alpha\gamma}{\gamma+\beta}\varepsilon$. From Eqs. (26)-(27), we observe that the departure peak period starts earlier with $\left(1-\frac{\alpha}{\gamma+\beta}\right)\varepsilon$.

The above results are all based on the assumption that the travel time of the last driver is free-flow travel time. However, the travel times of the last driver in the CPI model and in the FTL model are both endogenously determined by the condition that the total number of drivers arriving at the destination should be equal to the travel demand. The next section will show that the last driver's travel time is numerically very close to the free-flow travel time. This means the resulting equilibrium generalized travel cost and equilibrium peak period will also be very close to those in Eqs. (26)-(28). In fact, the differences turn out to be almost negligible, and therefore policy analysts can use the above closed-form results to get quick and fairly precise estimations of the equilibrium generalized travel costs, as well as the equilibrium peak period.

Section 3. Comparing models: numerical results

In the following, we use a numerical example to illustrate the models presented above, and will access the divergences caused by the alternative traffic density and travel time assumptions for the first and the last drivers.

Example 1: This example is used to illustrate the CPI model and the FTL model given in last section. The value of time and the shadow prices for early and late arrivals follows that in Verhoef (2005): $\alpha=7.5$, $\beta=3.75$, $\gamma=15$. The total travel demand is set to be $N=10000$, while the capacity of the bottleneck is $c_a=5000$ so that the minimum time needed for all drivers to pass the bottleneck is two hours. The length of the trip in the bathtub area is assumed to be $L=10$, and the free-flow traffic speed is $v_f=40$. Let $\lambda=0.00012$, so that the jam density is about $k_j=8333.3$. If jam density is reached with a vehicle every 5 meters, this corresponds to 41.7 km's of lanes, meaning that the average car travels quarter of the total lane capacity in the bathtub. The extra travel time above the free-flow travel time given in Eqs. (26)-(27) is $\varepsilon=0.0563$. Drivers' preferred arrival time at the destination is set to be $t^*=9$. The ratio between the flow of the bathtub and the inflow to the bottleneck is assumed to be equal to $\eta = 0.07982$, which is chosen in such a way that the maximum density in the CPI model is the same as in the FTL model.

Table 1 presents the numerical results of the CPI model and the FTL model. As benchmark models we use the closed-form approximations of section 2.6 assuming speed-density calculated travel time (SDTT) for the first driver while the last driver's travel time is free-flow travel time (FFTT). It can be observed that the divergences caused by the different assumptions are very small. The generalized travel costs for the two models increase by 0.30%, and 0.11%, respectively, compared to that in the approximate closed-form solution. Moreover, the generalized travel cost in the FTL model is closer to the benchmark than that in the CPI model. Similar conclusions can be obtained for the equilibrium peak period. In fact, compared to the benchmark, the equilibrium peak periods in both models are merely earlier/later less than 1 minute. Therefore, given the different assumptions about the relationship between the outflow and flow (e.g., the NEF in Gonzales & Daganzo, 2012) or the relationship between the travel time and traffic speed (e.g., Fosgerau, 2015), no substantial quantitative differences result for the resulting equilibrium generalized travel costs and peak periods.

For the traffic situation in the bathtub, we observe that the maximum equilibrium densities are only slightly different for both models. Both models have a hypercongested period in the bathtub that lasts about 16 minutes. The last driver's travel time is also affected by the traffic situation in the bathtub. We also find that the last driver' travel times in both cases are slightly higher than the free-flow travel time (less than 2 minutes). This means that when the last driver travels in the bathtub, the traffic speed is already very close to free-flow speed.

Table 1. Numerical results of the CPI model and the FTL model

Closed-form solution: the travel time of the last driver is free-flow travel time (section 2.6)	
Travel time of the first driver	SDTT (conventional bottleneck cost plus Speed-Density based Travel Time)
Departure period from origin	[7.1162, 9.1725]
Arrival period at destination	[7.4225, 9.4225]
Generalized travel cost	8.2127
Numerical: the travel time of the last driver is not free-flow travel time	

Travel time of the first driver	Method	CPI	FTL
	Value	0.3063	0.3063
Departure period from origin		[7.1096, 9.1492]	[7.1137, 9.1638]
Arrival period at destination		[7.4158, 9.4158]	[7.4200, 9.4200]
Travel time of the last driver		0.2666	0.2562
Generalized travel cost		8.2377	8.2221
Maximum density	Value	4744.0	4744.0
	Time	7.9016	7.9037
The critical density value for hypercongestion: 4166.7			
Hypercongestion period		[7.7316, 8.0139]	[7.7281, 8.0014]
Trip length		Varied	Fixed
Ratio between arrival rate at bottleneck and flow in bathtub		0.07982	Varied

To better understand the traffic situation in the bathtub, graphical representations of the various relevant traffic variables are given in this section (Fig 3(a, b)). We also give the equilibrium cumulative departures and arrivals (Fig 3(a, b)), and the resulting equilibrium travel time in the bathtub (Fig 4). First, Fig 3(a) and Fig 3(b) give the equilibrium cumulative departures and arrivals, traffic density, speed and flow in the bathtub over the peak period for the CPI model and the FTL model, respectively. When we compare Fig 3(a) and Fig 3(b), the shapes of the cumulative arrivals at the bottleneck in the CPI model and the FTL model are somewhat different. In Fig 3(a), the curve of the cumulative arrivals at the bottleneck is approximately linear at the early part of the peak period. Moreover, no queue exists in the last few minutes of the peak period in the CPI model, and drivers can pass the bottleneck as soon as they arrive. In contrast, the curve of the cumulative arrivals at the bottleneck in Fig 3(b) is clearly convex at the early part of the peak period and is approximately linear for the late part of the peak period.

Interestingly, the traffic variables in the bathtub are quite different in the late part of the peak period for the two model formulations. When we compare Fig 3(a) and Fig 3(b), we observe that the curve of the traffic density between $\tilde{t}$ and t_e is convex in the CPI model ($\left[7.9016, 9.1492\right]$) while it is first concave and then convex in the FTL model ($\left[7.9037, 9.1638\right]$). In particular, for the FTL model (Fig 3(b)), the traffic density decreases rather sharply after $\tilde{t}$, reaching a relatively low value in a short time. Then the curve of the traffic density becomes very flat for a long time. As a result, the corresponding traffic speed (flow) in the FTL model also rises (declines) very sharply over a short time and then stays flat over a substantially long time. This can also be verified by the horizontal distances between the cumulative departures from home and the cumulative arrivals at the bottleneck, which are very small and only have a minor decrease for the late part of the peak period. In contrast, the traffic density decreases gradually after $\tilde{t}$ until t_e in the CPI model (Fig 3(a)). As a result, the horizontal distances between the cumulative departures from home and the cumulative arrivals at the bottleneck are relatively large and decrease over time.

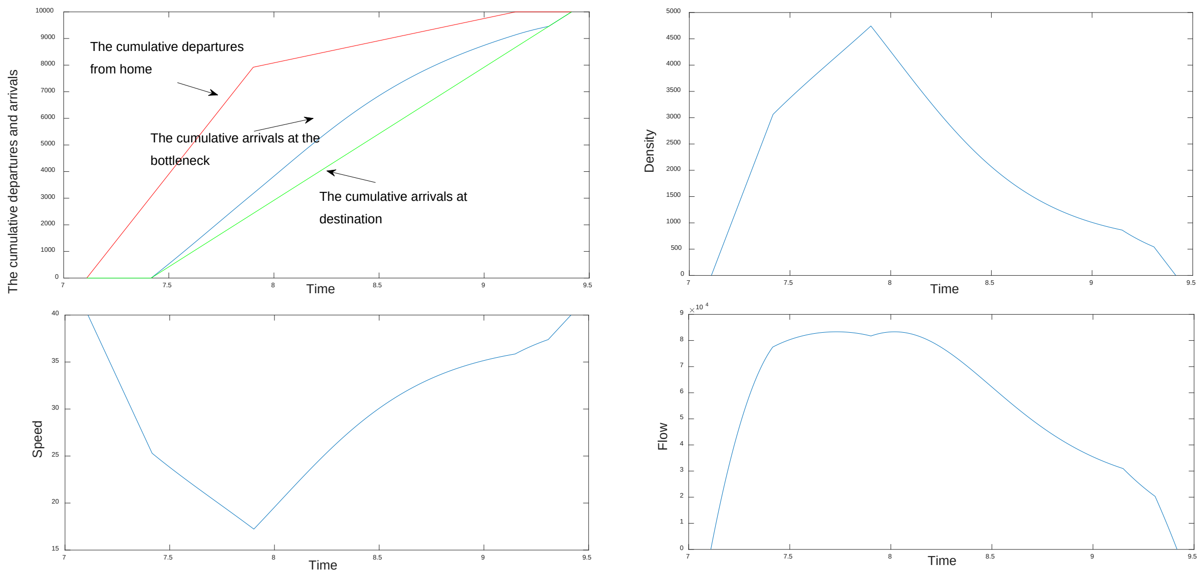


Fig 3(a). Equilibrium departure/arrival pattern and the traffic variables in the bathtub for the CPI model

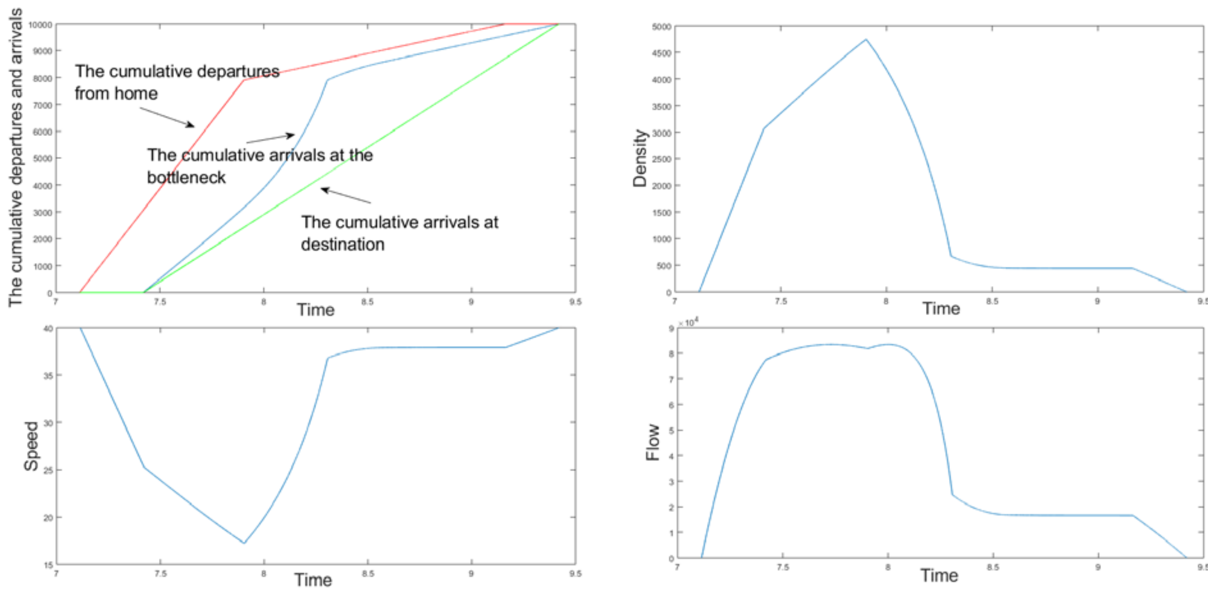


Fig 3(b). Equilibrium departure/arrival pattern and the traffic variables in the bathtub for the FTL model

Fig. 4 shows drivers' equilibrium travel times in the bathtub for the two approaches. The CPI model shows that drivers' equilibrium travel times in the bathtub both increase and decrease over a relatively long period. In contrast, the FTL model shows that drivers' equilibrium travel time in the bathtub increases to its maximal value in a relatively short time and is flat over an extended period in the late part of the peak period. As a result, the drivers departing at $\tilde{t}$ (at 7.9016) have a maximum travel time in the bathtub in the CPI model, while the maximum travel time in the bathtub occurs for drivers departing much earlier than $\tilde{t}$ (around 7.6). Moreover, the maximum travel time in the bathtub is higher for the CPI model compared to the FTL model. This is because for the CPI model, the actual trip length is not constant. This means that drivers may have a longer trip length, and thus a longer trip time.

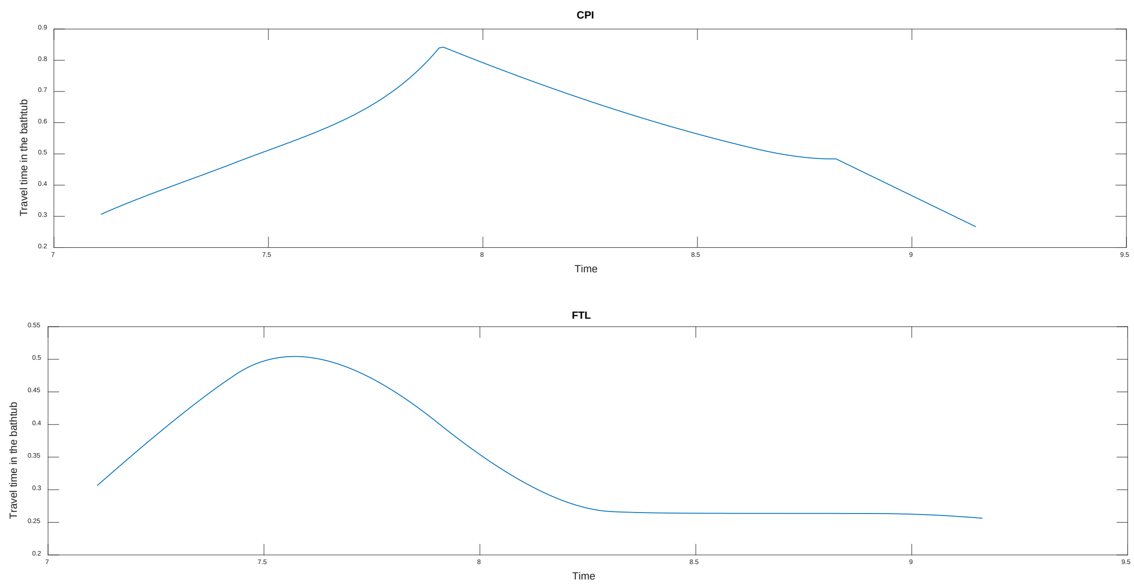


Fig 4. Drivers' travel time in the bathtub for the CPI model and the FTL model

Section 4. Hypercongestion

4.1 Two types of hypercongestion

This section discusses the qualitative insights on hypercongestion that we can derive from our model. Hypercongestion refers to the situation where the equilibrium traffic flow decreases with the equilibrium traffic density, because the impact of reduced speed dominates the direct effect of increased density on flow. For our model, this implies that one way to identify hypercongestion is to identify conditions where equilibrium traffic speed increases with the equilibrium traffic flow in the bathtub. This is fully in line with the empirical findings of Geroliminis & Daganzo (2008), who established a backward-bending part of the macroscopic speed-flow curve (the part in red oval in Fig 5(a)). As shown in Table 1 and Fig 3(a, b), hypercongestion of this type exists in the bathtub for the CPI and the FTL model. This type of hypercongestion affects travelers when they are in the bathtub area. However, it should be noted that for these equilibria, the bottleneck operates at full capacity for the whole hypercongestion period. The lower speed in the bathtub makes travelers join the bottleneck queue later, but since the outflow from the bottleneck is not affected, the arrival time at the destination does not change, either. Both the inflow into the bathtub and the outflow into the bottleneck queue are therefore unaffected by this hypercongestion in the bathtub, and although hypercongestion exists in the bathtub area, the performance of the full system of the bathtub-bottleneck in terms of arrival rates at the final destination is not affected. This happens because there is perfect substitution between the two types of travel delays: an increase of travel time in the bathtub leads to a shrinking queuing length at the bottleneck, and thus a decrease of queuing time perfectly substitutes the increase of travel time in the bathtub. Indeed, in dynamic equilibrium, it is the sum of the two types of delay that compensates for dynamic variation in schedule delay costs, so this sum will not change when hypercongested speeds occur in the bathtub.

However, there is a second type of hypercongestion that can occur: the increase of travel time in the bathtub may be so large, and the drop in the outflow from the bathtub into the bottleneck can be so strong, that the arrival rate at the bottleneck at some moments falls below the exit capacity, and if it does so for a sufficiently long time, the bottleneck may no longer be used at full capacity throughout the peak (see Fig 5(c)). For this type of hypercongestion, inefficiency caused by hypercongestion in the bathtub will affect the full system's arrival rate capacity. As a result, the equilibrium duration of the peak increases, and so does the equilibrium user cost (recall that the average user cost is constant over time in a dynamic equilibrium, so that a longer peak duration, with higher schedule delay cost for the very first traveler who faces no travel delay, implies a higher average cost level fir all drivers).

In single-facility models, these two types of hypercongestion coincide: a speed below the one that maximizes the flow, implies that the facility operates below its capacity. In our setting, in contrast, hypercongested speeds in the bathtub may or may not imply an arrival rate below the bottleneck's capacity for part of the peak. It is therefore important to distinguish these two types of hypercongestion. We call the former type of hypercongestion "bathtub-speed hypercongestion", as it only involves change of flow-speed relationship in the bathtub; and call the latter type of hypercongestion "system hypercongestion", because it causes an inefficiency in the performance of the full bathtub-bottleneck system.

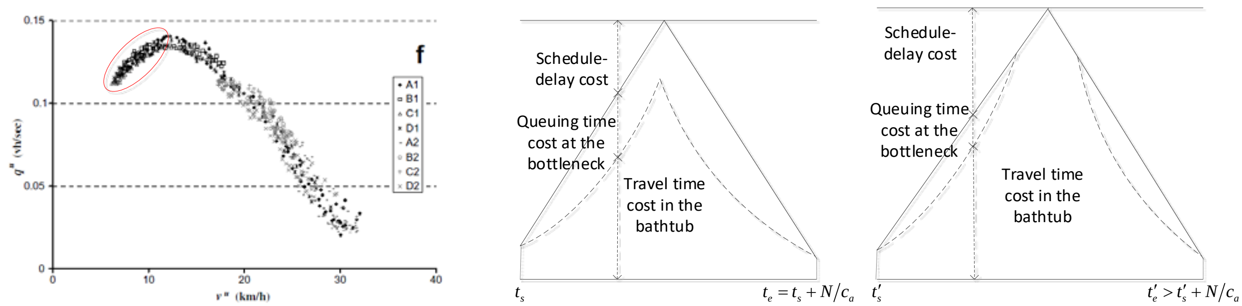


Fig 5(a). Average flow vs. average speed (Geroliminis and Daganzo, 2008)

Fig 5(b). The equilibrium user cost in the bathtub-speed hypercongestion

Fig 5(c). The equilibrium user cost in the system hypercongestion

Fig 5. Bathtub-speed hypercongestion and system hypercongestion

To fulfill the prior assumption that there will be no capacity waste at the bottleneck, parameters were chosen such that system hypercongestion does not occur in the numerical examples of Section 3. This underlines the intuitive notion that bathtub-speed hypercongestion is a necessary but not a sufficient condition for system hypercongestion to occur. However, if the capacity of the bathtub (represented by the jam density, $k_j = 1/\lambda$) further decreases, the assumption of no capacity waste at the bottleneck will not hold anymore, and for some drivers equilibrium travel times in the bathtub are so long that the arrival rates at the bottleneck are less than the exit capacity for some period. Therefore, system hypercongestion may arise, and *will* occur if the period of low arrival rates is long enough to make the queues disappear in the central period in the peak, so that outflow from the bottleneck falls below capacity. To investigate this kind of hypercongestion, the assumptions on parameters to secure that no capacity waste at the bottleneck exists will be changed in Section 4.3. But first, the next section analyzes bathtub-speed hypercongestion.

4.2 Bathtub-speed hypercongestion

Table 1 shows that bathtub-speed hypercongestion exists in both models. Based on the speed-density relationship in Eq. (1), the critical value of density for bathtub-speed hypercongestion is 4166.7. In the numerical results in Section 3, the maximum density occurs at $\tilde{t}$. The FTL model gives a maximum density (5469.5) that is much higher than this critical density value, and the duration of bathtub-speed hypercongestion is considerable (about 30 minutes). The maximum density is lower and bathtub-speed hypercongestion only exists during a short period (only a few minutes) in the CPI model.

Whether bathtub-speed hypercongestion exists depends on the values of various parameters, such as the total number of drivers, the preference parameters, and η. In order to show the impact of η on the occurrence of bathtub-speed hypercongestion, Fig 6 shows the equilibrium traffic density and flow in the CPI model for different values of η. The figure confirms the intuition that bathtub-speed hypercongestion is more likely to occur when η is smaller ($\eta < 1$); i.e., when the capacity of the exit from the bathtub into the bottleneck queue is smaller. Only then ($\eta = 0.07, 0.08, 0.09$) we observe that equilibrium density exceeds the critical value of 4166.7 in the upper left panel in combination with a drop in flow in the central peak in the upper right panel and longer equilibrium travel times in the lower left panel as a result. Consequently, the generalized user cost decrease with η (lower right panel).

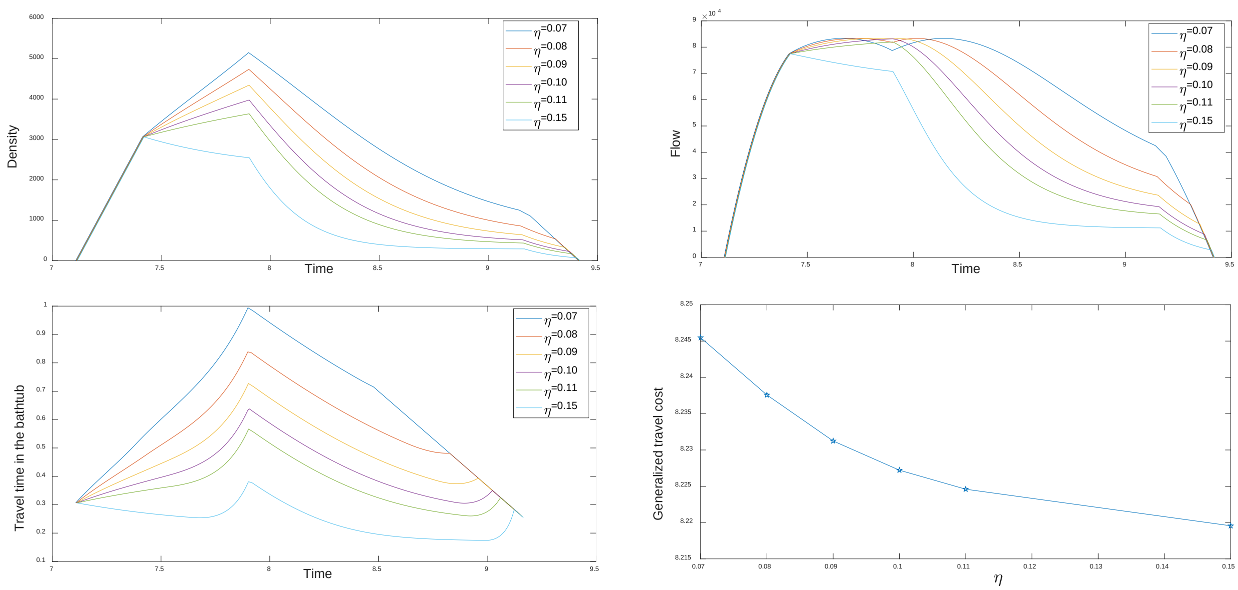


Fig 6. The impact of η in CPI model

The value of λ also plays an important role in the existence of hypercongestion, as it is the reciprocal of the jam density. Fig 7(a) and Fig 7(b) give the equilibrium traffic density, flow and travel time in the bathtub and the equilibrium generalized travel cost with different values of λ in the CPI model and the FTL model, respectively. The value of λ greatly affects the existence and the duration of hypercongestion in both models. The impact for the FTL model is more pronounced than for the CPI model, but the qualitative effects in both models are consistent. Traffic density during the whole peak period is higher when the value of λ increases (and hence the jam density declines). Equilibrium travel times in the bathtub therefore also increase with λ. The start time of the peak period is naturally earlier when λ increases and therefore the generalized travel cost also increases with λ.

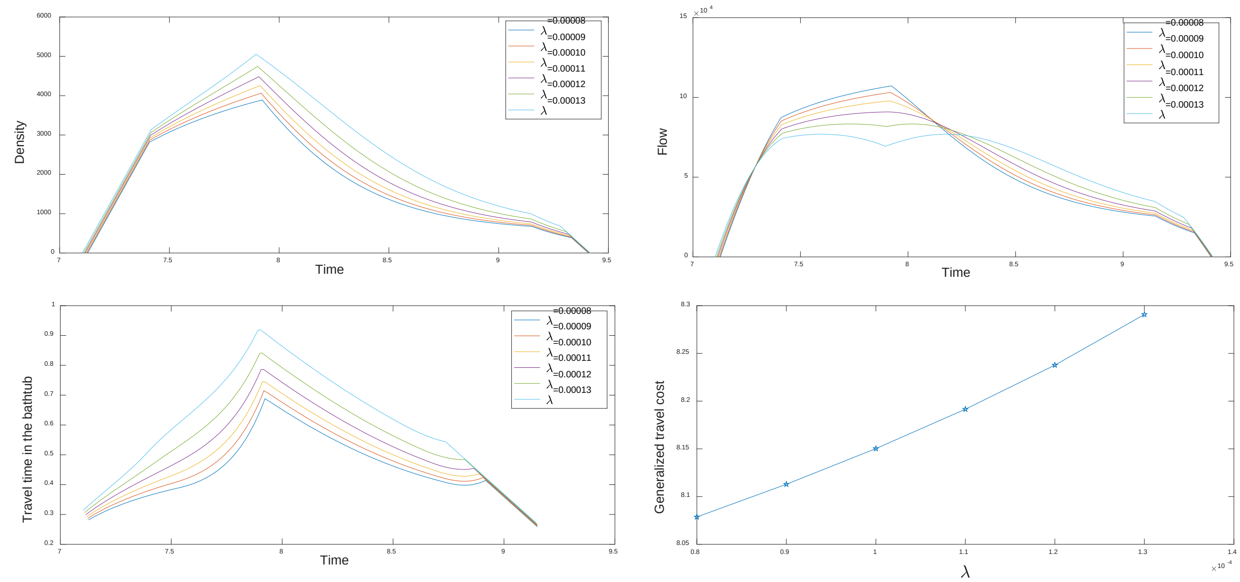


Fig 7(a). The CPI model with different values of λ

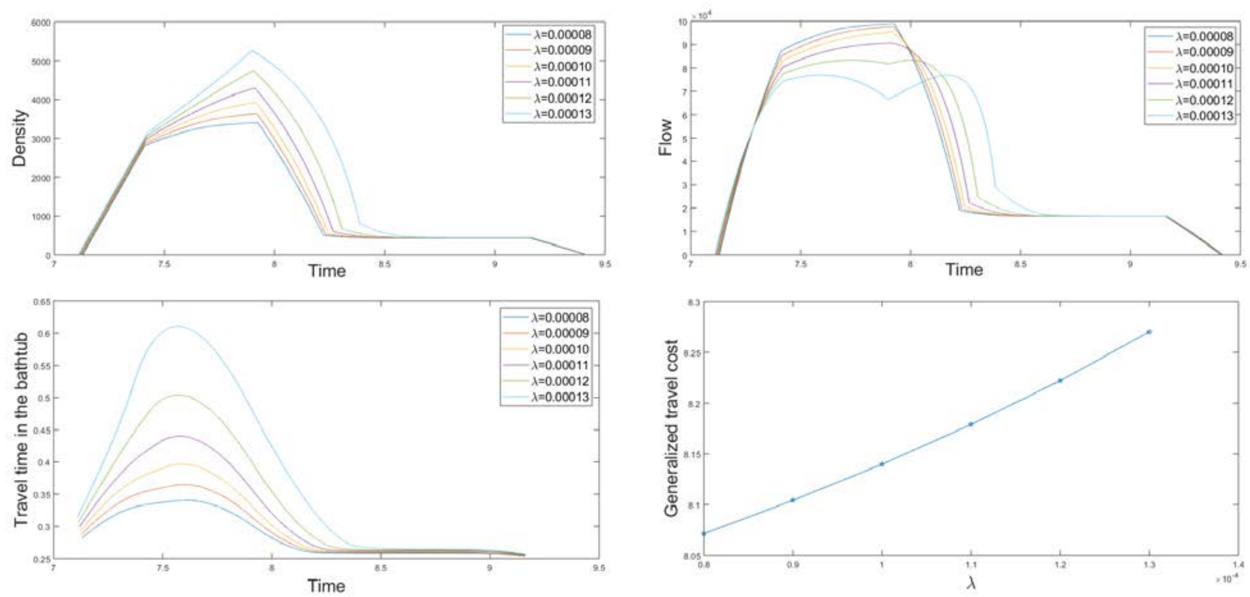


Fig 7(b). The FTL model with different values of λ

4.3 System hypercongestion

This section explores what happens when bathtub-speed hypercongestion lasts long enough to make the bottleneck queues disappear for some time, resulting in a capacity waste at the bottleneck in the middle of the peak period. As shown in Section 3, the differences between the equilibrium user cost and peak period in the CPI model and the FTL model are negligible and we will therefore only consider the CPI model in this section, as closed-form results of the traffic density in the bathtub exist (more details can be found in Appendix A). Denote $\bar{t}$ and $\hat{t}$ as the departure times of the first driver and the last driver who pass the bottleneck without queue in the central part of the peak, respectively. For drivers departing at $t \in \left[\bar{t}, \hat{t}\right]$, their arrival time at the bottleneck, and therefore also at the destination as no queue exists for them, is given by:

$$s(t) = \begin{cases} \frac{\alpha}{\alpha-\beta}\left(t-t_s\right)+t_q, \text{if } t_s \le t \le \tilde{t} \\ \frac{\alpha}{\alpha+\gamma}\left(t-\tilde{t}\right)+t^*, \text{ if } \tilde{t} < t \le t_e \end{cases}, t \in \left[\bar{t}, \hat{t}\right]. \tag{29}$$

However, the departure rate from home during this period does not satisfy Eq. (12) for $t \in \left[\bar{t}, \hat{t}\right]$, and the "constant proportion of inflow into the bottleneck", i.e., condition (17), does not hold for $t \in \left[s\left(\bar{t}\right), s\left(\hat{t}\right)\right]$. Therefore, for $t \in \left[\bar{t}, \hat{t}\right]$, drivers' travel time in the bathtub is given by:

$$s(t)-t = \begin{cases} \frac{\beta}{\alpha-\beta}t - \frac{\alpha}{\alpha-\beta}t_s + t_q, \text{if } t_s \le t \le \tilde{t} \\ t^* - \frac{\gamma}{\alpha+\gamma}t - \frac{\alpha}{\alpha+\gamma}\tilde{t}, \text{ if } \tilde{t} < t \le t_e \end{cases}, t \in \left[\bar{t}, \hat{t}\right], \tag{30}$$

and conditions (17)-(18) hold but the departure rate is unknown; for $t \in \left[s\left(\bar{t}\right), s\left(\hat{t}\right)\right]$, the departure rate is given by Eq. (12) but conditions (17)-(18) do not hold. Without further assumptions on the departure rate, the traffic density and speed in the bathtub for $t \in \left[\bar{t}, \hat{t}\right]$ cannot be obtained by solving Eq. (18). In fact, multiple solutions of the departure rate, and thus different traffic speeds in the bathtub during $\left[\bar{t}, \hat{t}\right]$ can lead to an equilibrium. This is because drivers' trip length is not required to be fixed, so as long as the travel time in the bathtub follows Eq. (16), different combinations of departure and traffic density satisfying condition (18) can lead to equilibrium (see also Fosgerau, 2015). For example, if we in addition assume that the trip length for drivers departing during $\left[\bar{t},\hat{t}\right]$ is fixed and equal to the average trip length L, based on condition (4), we have:

$$s'(t)=\frac{v\left(k(t)\right)}{v\left(k\left(s(t)\right)\right)}=\frac{1-\lambda\left(D(t)-A(t)\right)}{1-\lambda\left(D\left(s(t)\right)-D(t)\right)},t\in\left[\bar{t},\hat{t}\right]. \tag{31}$$

Here we only give a special case that $\hat{t}\le\tilde{t}$ and $s(\bar{t})\ge\tilde{t}$. As conditions (17)-(18) hold for $\left[\bar{t},\hat{t}\right]$, we can get $A(t)$, and as the departure rate follows Eq. (12) for $\left[s(\bar{t}),s(\hat{t})\right]$, we can get $D\left(s(t)\right)$. Combining Eq. (29) with the above Eq. (31) then yields:

$$D(t)=\frac{-\frac{\beta}{\lambda}+\alpha\left(\phi_2\left(s(t)-t_s\right)+C\right)+(\alpha-\beta)A(t)}{2\alpha-\beta}, \tag{32}$$

and hence:

$$d(t)=\frac{\alpha}{2\alpha-\beta}\frac{\alpha}{\alpha-\beta}\phi_2+\frac{\alpha-\beta}{2\alpha-\beta}a(t). \tag{33}$$

Substituting the above Eq. (33) into conditions (17)-(18) yields

$$k'(t)=\frac{\alpha}{2\alpha-\beta}\frac{\alpha}{\alpha-\beta}\phi_2-\frac{\alpha}{2\alpha-\beta}\eta k(t)v\left(k(t)\right). \tag{34}$$

Solving this differential equation gives the equilibrium traffic density in the bathtub for $t\in\left[\bar{t},\hat{t}\right]$. But other assumptions on L during $\left[\bar{t},\hat{t}\right]$ can also be consistent with equilibrium (although a different one in terms of departure patterns).

Fig 8 presents the CPI model with system hypercongestion, and hence capacity waste, in the middle of the peak. We set $\lambda=0.000145$, i.e., the jam density is equal to $k_j=6896.6$, while the values of the other parameters follow those in Section 3. Fig 8(a) confirms that the cumulative arrivals at the bottleneck between $\left[s(\bar{t}), s(\hat{t})\right]$ are equal to the cumulative arrivals at the destination, and the arrival rate is less than the capacity of the bottleneck. At the same time, the departure rate from home between $\left[\bar{t}, \hat{t}\right]$ follows Eq. (33), and is less than the departure rate at other moments in the early departure period (before $\bar{t}$ or after $\hat{t}$), which implies fewer drivers departing between $\left[\bar{t}, \hat{t}\right]$ to avoid the jam in the bathtub. At the same time, Fig 8(b) confirms that the equilibrium generalized travel cost during the peak period is identical for all drivers departing between the entire equilibrium departure period $[t_s, t_e]$, and would be higher outside. The duration of the equilibrium departure period (2.1520 hours in this case) is longer than without capacity waste (2 hours), and the equilibrium generalized travel cost (8.6911) is higher (increased by 5.50%). Consistent with this, the total travel time cost is higher (55390 vs 52365), and so is the total schedule delay cost (31521 vs 30012), underlining the inefficiency of system hypercongestion. Fig 8(c) and Fig 8(d) show the corresponding traffic density and the traffic flow in the bathtub. The whole hypercongestion period with bathtub-speed hypercongestion, is given by $\hat{t}$ to $s(\hat{t})$. However, system hypercongestion only exists between $\left[s(\bar{t}), s(\hat{t})\right]$ when the queuing time is zero.

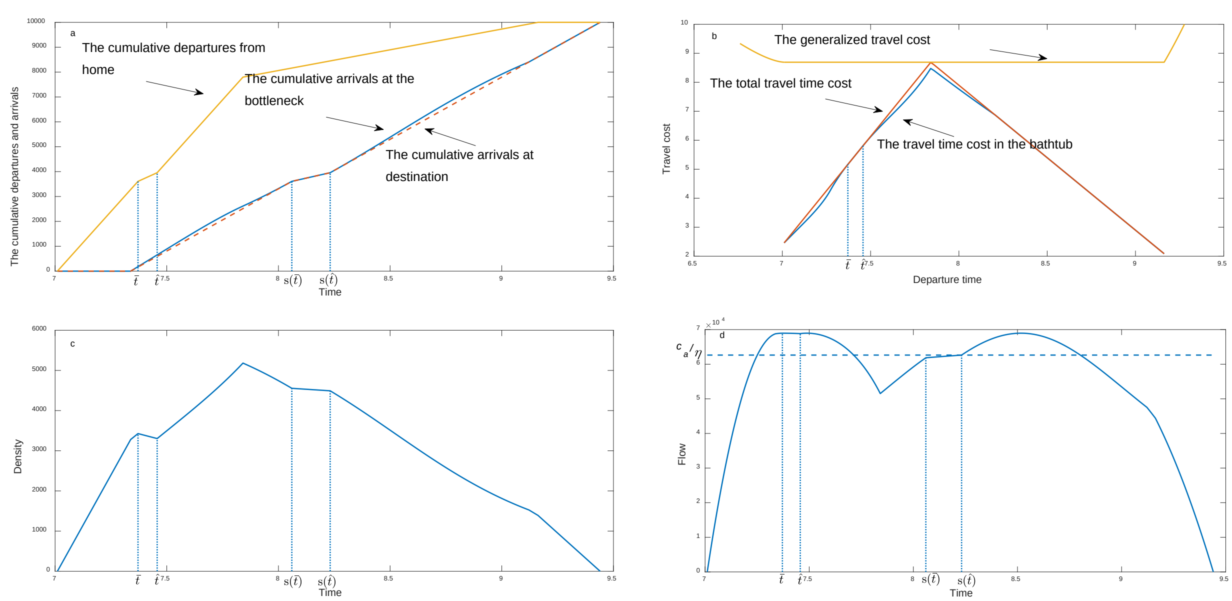


Fig 8. The CPI model with capacity waste in the middle of the peak period

In summary, the numerical results of Section 3 and in this section demonstrate that bathtub-speed hypercongestion can exist without the loss of efficiency, as reflected by outflow, of the whole system. This happens when traffic flow decreases with traffic density in the bathtub, but the bottleneck remains fully utilized. In this case, the drivers' total travel time does not change and it is merely a tradeoff between longer travel time in the bathtub and longer queuing time at the bottleneck. On the other hand, if the capacity of the bathtub (represented by the jam density) further decreases (or, the capacity of the bottleneck increases), system hypercongestion can occur. While traffic flow decreases with traffic density in the bathtub, the bottleneck is then not fully utilized resulting in an efficiency loss at the level of the full system.

5. Conclusion

This paper investigated the impacts of having a bathtub with a restricted exit capacity on equilibrium outcomes and user costs, using two model formulations from the literature. Different from earlier results one-directional facilities like in Verhoef (2003) we find that hypercongestion does not dissolve if the downstream capacity is increased, when the exit capacity is already higher than the equilibrium outflow from the bathtub. The reason is that spatial homogeneity of traffic conditions in the bathtub is imposed by construction, and no shockwave in upstream direction is possible in the bathtub, blocking the possible dissolution of hypercongestion from increased exit capacity by assumption.

Hypercongestion in the bathtub does not necessary lead to efficiency losses at the level of the full system as arrival flows at the final destination might not change. Two types of hypercongestion can thus be distinguished, bathtub-speed hypercongestion and system hypercongestion. When there is bathtub-speed hypercongestion, but the bottleneck remains fully utilized, longer travel times in the bathtub caused by hypercongested speeds result in shorter equilibrium queuing times at the exit and therefore the total equilibrium travel time of the trip does not change. On the contrary, when the drop in equilibrium bathtub flows is sufficiently large and long-lasting, the bottleneck may no longer be fully used throughout the most central peak period. Hypercongestion in the bathtub then results in a longer equilibrium peak period, and affects the full system's performance and thereby efficiency.

It was demonstrated in the numerical examples that bathtub-speed hypercongestion is a necessary but not a sufficient condition for system hypercongestion to occur. This implies that with a restricted exit capacity, whether it is necessary to regulate hypercongestion in an urban area depends on the performance of the exit capacity. Expanding the downstream capacity may be a better way to deal with hypercongestion than expanding flow capacity in the bathtub; a finding which is consistent with the one-directional model of Verhoef (2003).

Consistent with other findings in the literature (e.g., Amirgholy & Gao, 2017), we find that the numerical differences in the equilibrium generalized travel cost and peak period for the two model formulations of Fosgerau (2015) and Arnott et al. (2013) are almost negligible, and compared to an approximate model where it is assumed that the last driver's travel time is free-flow travel time are particularly negligible. The resulting analytical closed-form expressions for the generalized travel costs and the peak period of our approximate model can give policy analysts a quick estimations of these two terms, without having to know the specific traffic situation in the bathtub and at the bottleneck.

However, the predictions of the traffic situation in the bathtub and the queue between the two models can differ substantially. In fact, one of the main differences between the numerical results of the CPI model and the FTL model lies in the tradeoff between the equilibrium travel time in the bathtub and the equilibrium queuing time at the bottleneck. For the CPI model, the travel time in the bathtub accounts for the majority of the travel time cost whereas in the FTL model, the queuing time at the bottleneck accounts for the majority of the travel time costs. However, the sum of the travel time in the bathtub and the queuing time at the bottleneck at every instant in the CPI model and the FTL model are similar, as well as the total travel cost. It is interesting to investigate empirically in future research which of the two models predicts urban traffic patterns better. For now, it suggests that observationally diverging predictions in travel conditions may be consistent with rather similar predictions of equilibrium travel costs.

References

Amirgholy, M., & Gao, H. O. (2017). Modeling the dynamics of congestion in large urban networks using the macroscopic fundamental diagram: User equilibrium, system optimum, and pricing strategies. *Transportation Research Part B: Methodological*, *104*, 215–237.

Arnott, R. (2013). A bathtub model of downtown traffic congestion. *Journal of Urban Economics*, *76*, 110–121.

Arnott, R., & Buli, J. (2018). Solving for equilibrium in the basic bathtub model. *Transportation Research Part B: Methodological*, *109*, 150–175.

Arnott, R., de Palma, A., & Lindsey, R. (1990). Economics of a bottleneck. *Journal of Urban Economics*, *27*(1), 111–130.

Arnott, R., de Palma, A., & Lindsey, R. (1993). A structural model of peak-period congestion: a traffic bottleneck with elastic demand. *The American Economic Review, 83*(1), 161–179.

Arnott, R., & Inci, E. (2006). An integrated model of downtown parking and traffic congestion. *Journal of Urban Economics, 60*(3), 418–442.

Baer, N., Boucherie, R. J., & van Ommeren, J. K. C. W. (2019). Threshold Queueing to Describe the Fundamental Diagram of Uninterrupted Traffic. *Transportation Science*, *53*(2), 585–596.

Chu, X. (1995). Endogenous trip scheduling: the Henderson approach reformulated and compared with the Vickrey approach. *Journal of Urban Economics*, *37*, 324–343.

Daganzo, C. F. (2008). Queue Spillovers in Transportation Networks with a Route Choice. *Transportation Science*, *32*(1), 3–11.

Daganzo, C. F., & Lehe, L. J. (2015). Distance-dependent congestion pricing for downtown zones. *Transportation Research Part B: Methodological*, *75*, 89–99.

Fosgerau, M. (2015). Congestion in the bathtub. *Economics of Transportation*, *4*(4), 241–255.

Fosgerau, M., & Small, K. A. (2013). Hypercongestion in downtown metropolis. *Journal of Urban Economics*, *76*(1), 122–134.

Geroliminis, N., & Levinson, D. M. (2009). Cordon pricing consistent with the physics of overcrowding. *Transportation and Traffic Theory 2009: Golden Jubilee*, (1995), 219-240\r740.

Geroliminis, N. (2015). Cruising-for-parking in congested cities with an MFD representation. *Economics* *of Transportation*, *4*(3), 156–165.

Geroliminis, N., & Daganzo, C. F. (2007). Macroscopic modeling of traffic in cities. *TRB 86th Annual Meeting*, (January), 07–0413.

Geroliminis, N., & Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, *42*(9), 759–770.

Geroliminis, N., & Levinson, D. M. (2009). Cordon Pricing Consistent with the Physics of Overcrowding. In *Transportation and Traffic Theory 2009: Golden Jubilee* (pp. 219–240).

Gonzales, E. J., & Daganzo, C. F. (2012). Morning commute with competing modes and distributed demand: User equilibrium, system optimum, and pricing. *Transportation Research Part B: Methodological*, *46*(10), 1519–1534.

Greenshields, B. D., Bibbins, J. R., Channing, W. S., & Miller, H. H. (1935). A study of traffic capacity. *Highway Research Board Proceedings. National Research Council (USA), Highway Research Board.*, *14*.

Henderson, J. V. (1974). Road Congestion A Reconsideration of Pricing Theory. *Journal of Urban Economics*, *1*(3), 346–365.

Lamotte, R., & Geroliminis, N. (2018). The morning commute in urban areas with heterogeneous trip lengths. *Transportation Research Part B: Methodological*, *117*, 794–810.

Liu, W., & Geroliminis, N. (2016). Modeling the morning commute for urban networks with cruising-for-parking: An MFD approach. *Transportation Research Part B: Methodological*, *93*, 470–494.

Ma, R., Ban, X. (Jeff), & Pang, J. S. (2017). A Link-Based Differential Complementarity System Formulation for Continuous-Time Dynamic User Equilibria with Queue Spillbacks. *Transportation Science*, *52*(3), 564–592.

Mariotte, G., & Leclercq, L. (2019). Flow exchanges in multi-reservoir systems with spillbacks. *Transportation Research Part B: Methodological*, *122*, 327–349.

Mariotte, G. , Leclerc, L. , Laval, J. (2017). Macroscopic urban dynamics: analytical and numerical comparisons of existing models. *Transportation Research Part B: Methodological*, 101, 245–267.

Mun, S. (1994). Traffic jams and the congestion. *Transportation Research Part B: Methodological*, *28B*(5), 365–375.

Mun, S. (1999). Peak-load pricing of a bottleneck with traffic jam. *Journal of Urban Economics*, *46*(3), 323–349.

Newell, G. F. (1988). Traffic Flow for the Morning Commute. *Transportation Science*, *22*(1), 47–58.

Nie, Y. (Marco), & Zhang, H. M. (2008). Oscillatory Traffic Flow Patterns Induced by Queue Spillback in a Simple Road Network. *Transportation Science*, *42*(2), 236–248.

Small, K. A., & Chu, X. (2003). Hypercongestion. *Journal of Transport Economics and Policy, 37*(3), 319–352.

Verhoef, E. T. (2001). An integrated dynamic model of road traffic congestion based on simple car-following theory: exploring hypercongestion. *Journal of Urban Economics*, *49*(3), 505–542.

Verhoef, E. T. (2003). Inside the queue: hypercongestion and road pricing in a continuous time–continuous place model of traffic congestion. *Journal of Urban Economics*, *54*(3), 531–565.

Verhoef, E. T. (2005). Second-best congestion pricing schemes in the monocentric city. *Journal of Urban Economics*, *58*(3), 367–388.

Vickrey, W. S. (1969). Congestion theory and transport investment. *American Economic Review (Papers and Proceedings)*, *59*(2), 251–260.

Yuan, K., Knoop, V. L., & Hoogendoorn, S. P. (2017). A Microscopic Investigation Into the Capacity Drop: Impacts of Longitudinal Behavior on the Queue Discharge Rate. *Transportation Science*, *51*(3), 852–862.

Acknowledgement

This work is supported by funding of the Netherlands Organization for Scientific Research (NWO) as part of the U-SMILE project 438-15-176, which is gratefully acknowledged.

Appendix A. The traffic density in the bathtub for the CPI mode

Here we assume $t_q \le \tilde{t}$, and the density function with $t_q > \tilde{t}$ can be obtained in a similar way. As discussed in the last section, traffic density in the bathtub is obtained by solving the differential equation Eq. (18), which is given by, if $\eta v_f - 4\lambda\phi_1 > 0$,

$$k(t) = \begin{cases} \dfrac{1-\rho_1}{2\lambda} + \dfrac{\rho_1}{\lambda\left(1+\exp\left(\varphi_1 t + c\right)\right)}, t_q \le t \le \tilde{t} \\ \dfrac{1-\rho_2}{2\lambda} + \dfrac{\rho_2}{\lambda\left(1+\exp\left(\varphi_2 t + c'\right)\right)}, \tilde{t} < t \le t_e \end{cases}, \tag{A1}$$

if $\eta v_f - 4\lambda\phi_1 = 0$,

$$k(t) = \begin{cases} \dfrac{1}{2\lambda} - \dfrac{1}{4\lambda^2\left(\phi_1 t + c\right)}, t_q \le t \le \tilde{t} \\ \dfrac{1-\rho_2}{2\lambda} + \dfrac{\rho_2}{\lambda\left(1+\exp\left(\varphi_2 t + c'\right)\right)}, \tilde{t} < t \le t_e \end{cases}, \tag{A2}$$

if $4\lambda\phi_2 \le \eta v_f < 4\lambda\phi_1$,

$$k(t) = \begin{cases} \dfrac{1}{2\lambda}\rho_1 \tan\left(\dfrac{1}{2}\varphi_1\left(t + c\right)\right) + \dfrac{1}{2\lambda}, t_q \le t \le \tilde{t} \\ \dfrac{1-\rho_2}{2\lambda} + \dfrac{\rho_2}{\lambda\left(1+\exp\left(\varphi_2 t + c'\right)\right)}, \tilde{t} < t \le t_e \end{cases}, \tag{A3}$$

where c and c' are parameters such that $k(t)$ is continuous at t_q and $\tilde{t}$, and for convenience we denote

$\rho_1 = \dfrac{\sqrt{\eta v_f \mp 4\lambda\phi_1}}{\sqrt{\eta v_f}}$, $\varphi_1 = \sqrt{\eta v_f\left(\eta v_f \mp 4\lambda\phi_1\right)}$, $\rho_2 = \dfrac{\sqrt{\eta v_f - 4\lambda\phi_2}}{\sqrt{\eta v_f}}$, $\varphi_2 = \sqrt{\eta v_f\left(\eta v_f - 4\lambda\phi_2\right)}$.

As there are no arrivals at the bottleneck before the first departed arrives there, the density is equal to the cumulative departures for $t_s \le t < t_q$, i.e.,

$$k(t) = D(t) = \phi_1\left(t - t_s\right), t_s < t \le t_q. \tag{A4}$$

Consistent with this, there are no departures after t_e while arrivals at the bottleneck still occur. Therefore, for $t_e < t \le s(t_e)$, we have:

$$k(t) = N - A(t) = N - \eta v_f \int_{t_q}^{t} k(\omega)\left(1-\lambda k(\omega)\right) d\omega, t_e < t \le s(t_e), \tag{A5}$$

which yields:

$$k(t) = \frac{1}{\exp\left(\eta v_f \left(t + c''\right)\right) + \lambda}, t_e < t \le s(t_e), \tag{A6}$$

where c'' is parameter such that $k(t)$ is continuous at t_e. Based on condition (2), the cumulative arrivals are equal to the difference between the cumulative departures and the density, $A(t) = D(t) - k(t)$. Therefore,

$$A\left(s(t_e)\right) = N - \frac{1}{\exp\left(\eta v_f \left(s(t_e) + c''\right)\right) + \lambda}, \tag{A7}$$

which implies the last driver departing will approach the bottleneck, but will never arrive. This is caused by the homogenous density assumption, as the last driver needs to circuit in the bathtub to keep homogenous density over space. It means that the peak period will last forever, which is impossible.

Fortunately, it also implies capacity waste may occur at the bottleneck for the very last few minutes of the peak period, which means the equilibrium condition (11) cannot apply. Denote $s(\hat{t})$ as the instant when the length of queue at the bottleneck becomes zero, namely the cumulative arrivals at the bottleneck is equal to the cumulative arrivals at destination:

$$A\left(s(\hat{t})\right) = c_a\left(s(\hat{t}) - t_q\right). \tag{A8}$$

For drivers departing after $\hat{t}$, they can pass the bottleneck as soon as they arrive at it. Thus, their generalized travel cost is given by:

$$u(t) = \alpha\left(s(t) - t\right) + \gamma\left(s(t) - t^*\right), t \in \left[\hat{t}, t_e\right]. \tag{A9}$$

The first-order condition of equilibrium is:

$$u'(t) = (\alpha + \gamma)s'(t) - \alpha = 0, t \in \left[\hat{t}, t_e\right], \tag{A10}$$

which yields:

$$s'(t)=\frac{\alpha}{\alpha+\gamma}, t\in[\hat{t},t_e]. \tag{A11}$$

From condition (5), the departure rate satisfies:

$$d(t)=a(s(t))s'(t)=a(s(t))\frac{\alpha}{\alpha+\gamma}, t\in[\hat{t},t_e]. \tag{A12}$$

As there is no queue at the bottleneck after $s(\hat{t})$, the arrival rate at the bottleneck, $a(s(t))$, should be no higher than the capacity of the bottleneck, which means the departure rate should be no higher than ϕ_2. For simplicity, here we further assume the departure rate between $[\hat{t},t_e]$ is still equal to ϕ_2, and thus, the arrival rate at the bottleneck is equal to the capacity of the bottleneck. In this way, the bottleneck is still fully used and no queue exists. Therefore, the cumulative departures still follow Eq. (13) between $[\hat{t},t_e]$, and the cumulative arrives at the bottleneck between $[s(\hat{t}),s(t_e)]$ is given by:

$$A(t)=c_a(t-t_q), t\in(s(\hat{t}),s(t_e)) \tag{A13}$$

Thus, the density between $[s(\hat{t}),s(t_e)]$ is given by:

$$k(t)=N-c_a(t-t_q), t\in(s(\hat{t}),s(t_e)) \tag{A14}$$

and all drivers arrive at their destination at $s(t_e)=t_q+N/c_a$. Although drivers can also choose lower departure rates in this period to reach equilibria, they all lead to longer equilibrium peak period and higher equilibrium travel cost. For peaks in which these light travel conditions in the very last instants are relatively short-lived compared to the full peak duration, and are of little importance because of their short duration and the low flow levels in those intervals, the assumption seems acceptable, definitely in the light of the analytical advantages it brings.

Appendix B. Alternative assumptions on traffic density at the shoulder of the peak period

Another interesting question is whether the assumption of uniformly distributed travelers in the bathtub should be maintained for the "shoulders" of the peak. Clearly, especially for the shoulders of the peak when Eq. (14) or Eq. (15) applies, the assumption is less appropriate as it means that the first driver, when entering the bathtub, is immediately “scattered over space”, in order to produce a spatially homogenous traffic density. This also would be the case for the last driver, just before leaving the bathtub. Both situations have the merit of maintaining consistency of assumptions, but the difficulty of being far from reality, while possibly having a strong impact on equilibrium outcomes.

In fact, if we further consider the situations of the first and the last driver, one could argue that in reality there is no traffic before (after) the first (last) driver, and therefore the vehicles in the bathtub should only occupy a part of the bathtub. In other words, when the first/last driver travels, the size of the area in the bathtub actually occupied with traffic becomes larger/smaller, defined by the location of the first/last driver departed. In this way, the size of the traffic area is proportional to the trip length of the first driver (i.e., $\int_{t_s}^{t} v\left(k(\omega)\right) d\omega$) (or the remaining trip length of the last driver, i.e., $L - \int_{t_e}^{t} v\left(\lambda k(\omega)\right) d\omega$) and in that way we may maintain uniform conditions within a shrinking part of the bathtub:

$$k(t) = \frac{L}{\int_{t_s}^{t} v\left(k(\omega)\right) d\omega} D(t), t_s < t \le t_q, \tag{B1}$$

$$k(t) = \frac{L}{L - \int_{t_e}^{t} v\left(\lambda k(\omega)\right) d\omega} \left(N - A(t)\right), t_e < t \le s(t_e). \tag{B2}$$

From Eqs. (B1)-(B2), the derivatives of density with respect to time in these two periods are given by:

$$k'(t) = \frac{L\phi_1 - F(t)}{\int_{t_s}^{t} v(\omega) d\omega}, t_s < t \le t_q, \tag{B3}$$

$$k'(t) = \frac{(1 - \eta L) F(t)}{L - \int_{t_e}^{t} v(\omega) d\omega}, t_e < t \le s(t_e). \tag{B4}$$

From Eq. (B3), when time t is very close to t_s, $\int_{t_s}^{t} v(\omega) d\omega$ is very close to zero and thus, $k'(t)$ is very close to infinite. This means the traffic density will increase sharply to a very high value within a short time after the first driver departs, which is clearly impossible. As to the period when the last driver travels, from Eq. (B4), $k'(t)$ is positive if $1 - \eta L > 0$, which means traffic density increases with time. This is only possible if the decrease of the bathtub size is faster than the decrease of the density in the bathtub. However, it may also result in infinite travel time for the last driver. On the other hand, if $1-\eta L<0$, $k'(t)$ is negative and thus, density decreases with time until zero, which implies that the decrease of the density in the bathtub is faster than the decrease of the bathtub size. However, this means $\eta > 1/L$. In other words, the capacity of the exit of the bathtub should be larger than the size of the bathtub. Even when equilibrium exists, the resulting numerical results (available from the authors on request) show that the different assumptions of traffic density at the extreme shoulders do not have significant impacts on the estimations of equilibrium generalized travel cost and the equilibrium traffic situation in the full peak period.